

Claims

1. A method for managing memory resources in a service gateway environment, the gateway environment having a plurality of service instances where at least one of the service instances being dependent on another service instance, comprising:

receiving a service request, the service request having an associated memory space requirement that exceeds total available memory space associated with the gateway environment;

determining a number of dependent service instances for each service instance;

determining an accumulative memory space requirement for each service instance, the accumulative memory space requirement accounting for service instances that depend from a given service instance;

identifying a subset of service instance whose memory space requirement exceeds the memory space requirement of the service request, the subset having a minimal number of the service instances; and

performing a memory resource management operation in relation to the identified subset of service instances.

2. The method of Claim 1 wherein the step of performing a memory resource management operation further comprises deleting the identified subset of service instances.

3. The method of Claim 1 wherein the step of identifying a subset of service instances further comprises determining a ratio for each service instance, where the ratio is defined as the accumulative memory space requirement divided by the number of dependent service instances, and selecting the service instance having the largest ratio and each of the service instances that depend from the service instance having the largest ratio.

4. The method of Claim 1 wherein the step of identifying a subset of service instances further comprises determining an order for traversing the plurality of service instances; building a dynamic programming table for the plurality of service instances, such that the entries in the table indicate an amount of memory space that can be attained by deleting a subset of the service instances; and identifying one or more service instances that are to be deleted using the dynamic programming table.

5. The method of Claim 1 wherein the step of determining an accumulative memory space requirement further comprises summing memory space requirements for a given active service instance with memory space

requirements for each of the active services instances that depend from the given active service instance

6. A method for managing memory resources in a service gateway environment, the gateway environment having a plurality of service instances, comprising:

(a) receiving a service request, the service request having an associated memory space requirement that exceeds total available memory space associated with the gateway environment;

(b) determining an accumulative memory space requirement for each service instance;

(c) determining a number of dependent service instances for each service instance;

(d) determining a ratio for each service instance, where the ratio is defined as the accumulative memory space requirement divided by the number of dependent service instances; and

(e) performing a memory resource management operation in relation to the service instance having the largest ratio.

7. The method of Claim 6 wherein the step of performing a memory resource management operation further comprises deleting the service instance having the largest ratio and each of the services instances that depend from the service instance having the largest ratio.

8. The method of Claim 7 further comprises:

determining the total available memory space associated with the gateway environment, after the step of deleting the service instance;

determining whether the memory space requirement of the service request exceeds the total available memory space associated with the gateway environment; and

repeating steps (b) thru (e) when the memory space requirement of the service request exceeds the total available memory space associated with the gateway environment.

9. The method of Claim 8 further comprises the step of performing the service request when the memory space requirement of the service request is less than or equal to the total available memory space associated with the gateway environment.

10. The method of Claim 6 wherein the step of determining an accumulative memory space requirement further comprises summing memory space requirements for a given service instance with memory space requirements for each of the services instances that depend from the given service instance.

11. The method of Claim 6 further comprises maintaining a data store for the plurality of service instances, wherein the data store includes an identifier for a given service instance, the accumulative memory space requirement for the given service instance, and the number of service instances that depend from the given service instance.

12. A method for managing memory resources in a service gateway environment, the gateway environment having a plurality of service instances, comprising:

receiving a service request, the service request having an associated memory space requirement that exceeds total available memory space associated with the gateway environment;

determining an accumulative memory space requirement for each service instance in the gateway environment;

determining an order for traversing the plurality of service instances;

building a dynamic programming table for the plurality of service instances, such that the entries in the table indicate an amount of memory space that can be attained by deleting a subset of the service instances; and

identifying one or more service instances that are to be deleted using the dynamic programming table.

13. The method of Claim 12 further comprises the step of deleting the identified service instances, thereby managing memory resources in the gateway environment.

14. The method of Claim 12 wherein the step of determining an accumulative memory space requirement further comprises summing memory space requirements for a given service instance with memory space requirements for each of the services instances that depend from the given service instance.

15. The method of Claim 12 wherein the step of determining an order for traversing further comprises recursively traversing service instances in a post-order.

16. The method of Claim 15 where the step of building a dynamic programming table further comprises defining each row in the table to correlate to a subset of the service instances in accordance with the post-order and defining each column in the table to correlate to a number of service instances that are to be deleted from the subset of service instances, such that each entry in the table indicates a maximum amount of memory space that can be attained by deleting the corresponding number of service instance from the corresponding subset of service instances.

17. The method of Claim 12 wherein the step of identifying one or more service instances that are to be deleted further comprises evaluating from left to right the entries of a bottom row of the dynamic programming table and selecting service instances that correspond to an entry whose memory space exceeds the memory space requirement of the service request.